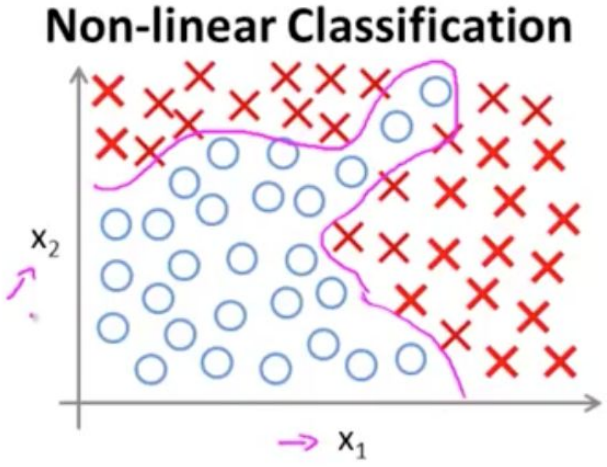
MACHINE LEARNING
NEURAL NETWORKS REPRESENTATION

Machine Learning – Andrew Ng

HueAI – April 2019

NON-LINEAR HYPOTHESES



$x_1 =$ size
$x_2 =$ # bedrooms
$x_3 =$ # floors
$x_4 =$ age
$\cdots$
x_{100}

$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \theta_6 x_1 x_2^2 + \cdots)$$

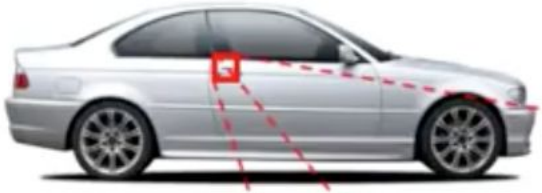
$$x_1^2, x_1 x_2, x_1 x_3, \ldots, x_1 x_{100}, x_2^2, x_2 x_3, \ldots \approx 5000\ features.$$

$$\approx \frac{n^2}{2}$$

$$x_1 x_2 x_3, x_1 x_3 x_4, \ldots, x_1 x_{99} x_{100}, x_2 x_3 x_4, \ldots \approx O(n^3). \qquad \approx 170.000$$

WHAT IS THIS?

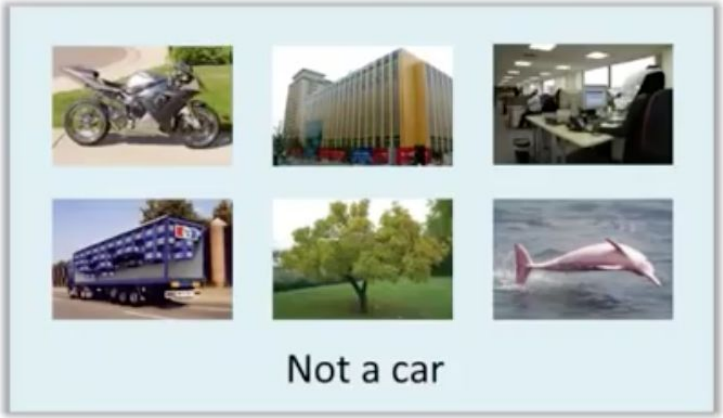
You see this:



But the camera sees this:

194	210	201	212	199	213	215	195	178	158	182	209
180	189	190	221	209	205	191	167	147	115	129	163
114	126	140	188	176	165	152	140	170	106	78	88
87	103	115	154	143	142	149	153	173	101	57	57
102	112	106	131	122	138	152	147	128	84	58	66
94	95	79	104	105	124	129	113	107	87	69	67
68	71	69	98	89	92	98	95	89	88	76	67
41	56	68	99	63	45	60	82	58	76	75	65
20	43	69	75	56	41	51	73	55	70	63	44
50	50	57	69	75	75	73	74	53	68	59	37
72	59	53	66	84	92	84	74	57	72	63	42
67	61	58	65	75	78	76	73	59	75	69	50

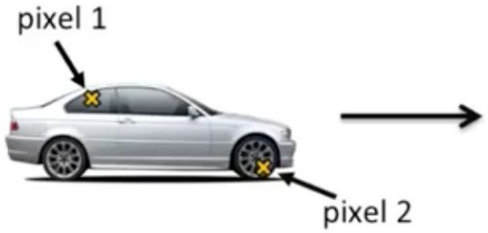
COMPUTER VISION
CAR DETECTION



Testing:

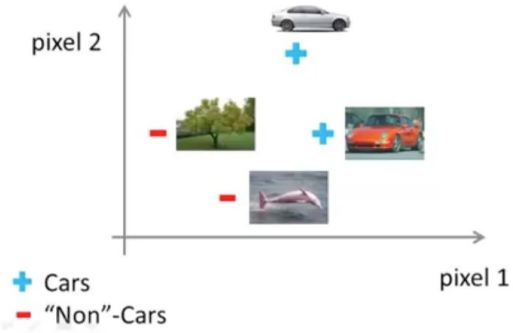
What is this?

CAR DETECTION

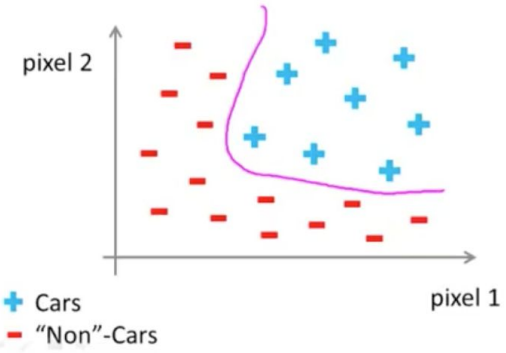


Learning Algorithm

$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \theta_6 x_1 x_2^2 + \cdots)$$

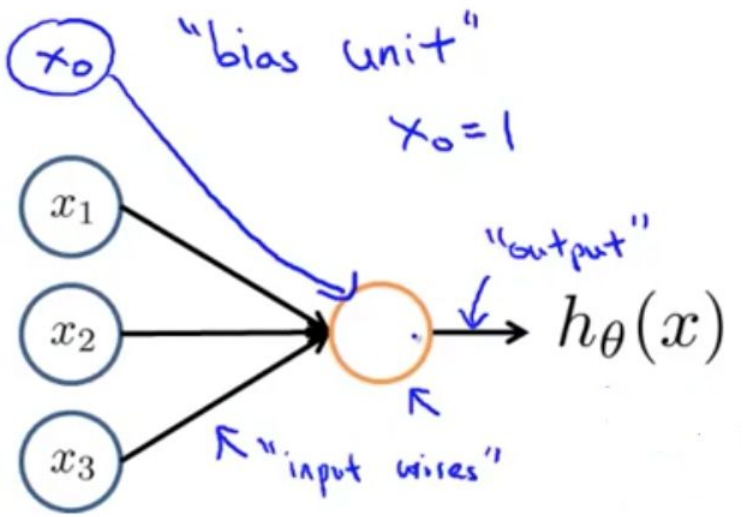


$$x_1^2, x_1 x_2, x_1 x_3, \ldots, x_1 x_{100}, x_2^2, x_2 x_3, \ldots$$
$$\approx 5000\ features.$$
$$\approx \frac{n^2}{2}$$



$$x_1 x_2 x_3, x_1 x_3 x_4, \ldots, x_1 x_{99} x_{100}, x_2 x_3 x_4, \ldots$$
$$\approx O(n^3). \qquad \approx 170.000$$

NEURON MODEL: LOGISTIC UNIT



$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \theta_6 x_1 x_2^2 + \cdots)$

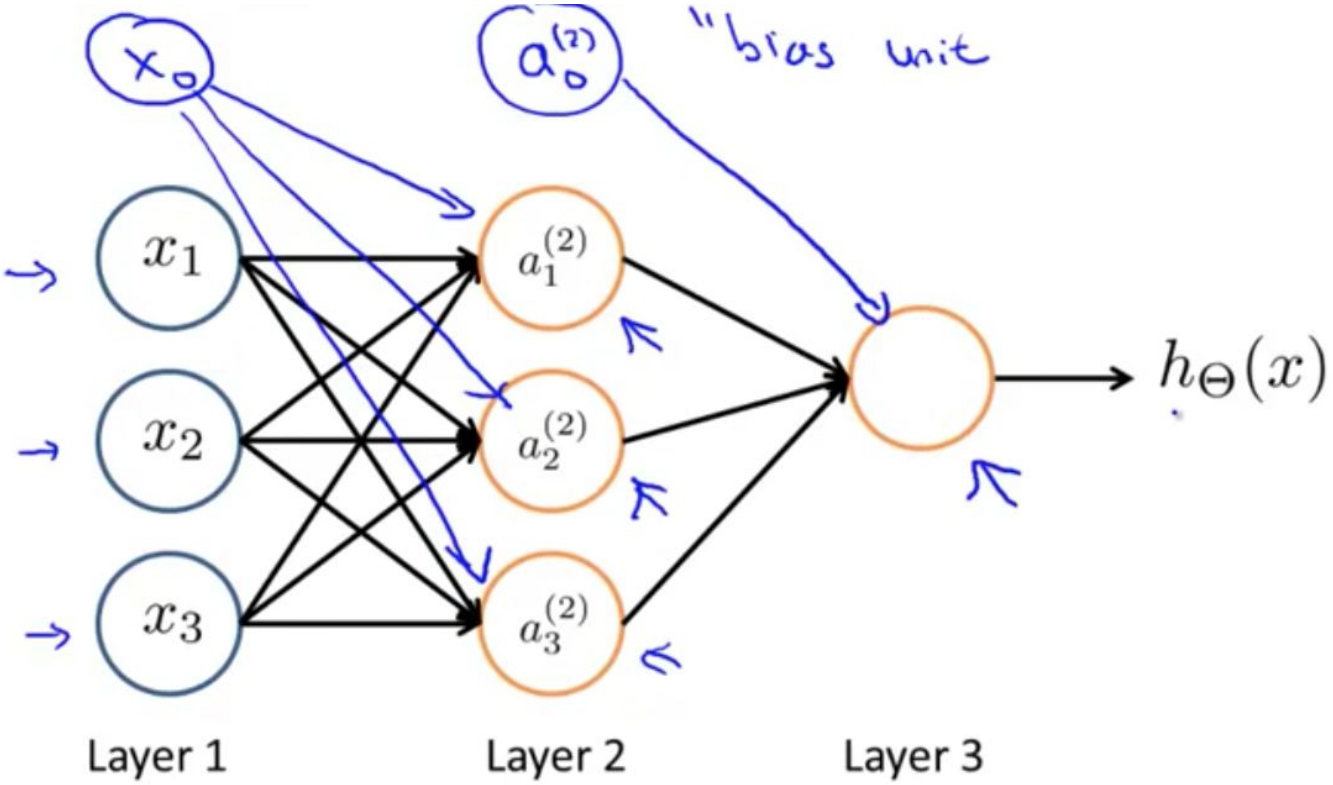
$x_1^2, x_1 x_2, x_1 x_3, \ldots, x_1 x_{100}, x_2^2, x_2 x_3, \ldots$
$\approx 5000\ features.$
$\approx \frac{n^2}{2}$

$x_1 x_2 x_3, x_1 x_3 x_4, \ldots, x_1 x_{99} x_{100}, x_2 x_3 x_4, \ldots$
$\approx O(n^3).$ ≈ 170.000

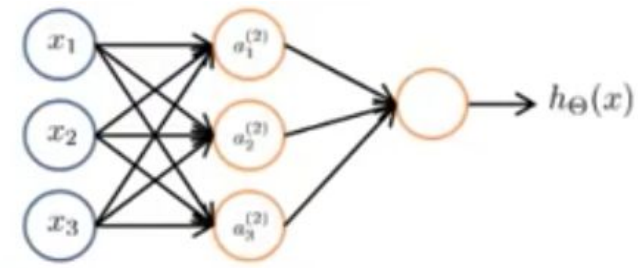
$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \theta_6 x_1 x_2^2 + \cdots)$

$x_1^2, x_1 x_2, x_1 x_3, \ldots, x_1 x_{100}, x_2^2, x_2 x_3, \ldots$
$\approx 5000\ features.$
$\approx \frac{n^2}{2}$

NEURAL NETWORK



NEURAL NETWORK



$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \theta_6 x_1 x_2^2 + \cdots)$$

$$x_1^2, x_1 x_2, x_1 x_3, \ldots, x_1 x_{100}, x_2^2, x_2 x_3, \ldots \approx 5000\ features. \approx \frac{n^2}{2}$$

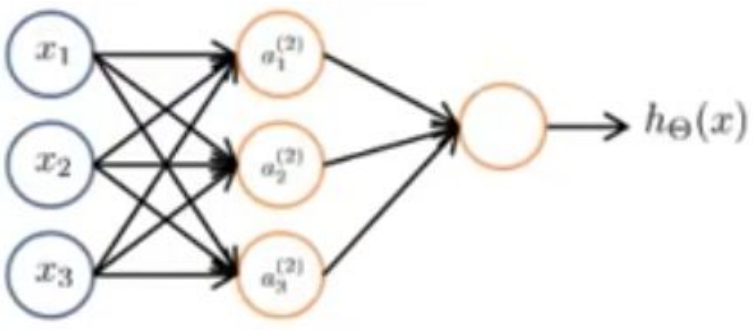
$$x_1 x_2 x_3, x_1 x_3 x_4, \ldots, x_1 x_{99} x_{100}, x_2 x_3 x_4, \ldots \approx O(n^3). \quad \approx 170.000$$

$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \theta_6 x_1 x_2^2 + \cdots)$$

$$x_1^2, x_1 x_2, x_1 x_3, \ldots, x_1 x_{100}, x_2^2, x_2 x_3, \ldots \approx 5000\ features. \approx \frac{n^2}{2}$$

$$x_1 x_2 x_3, x_1 x_3 x_4, \ldots, x_1 x_{99} x_{100}, x_2 x_3 x_4, \ldots \approx O(n^3). \quad \approx 170.000$$

FORWARD PROPAGATION: VECTORIZED IMPLEMENTATION



$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \theta_6 x_1 x_2^2 + \cdots)$$

$$x_1^2, x_1 x_2, x_1 x_3, \ldots, x_1 x_{100}, x_2^2, x_2 x_3, \ldots \approx 5000\ features.$$

$$\approx \frac{n^2}{2}$$

$$x_1 x_2 x_3, x_1 x_3 x_4, \ldots, x_1 x_{99} x_{100}, x_2 x_3 x_4, \ldots \approx O(n^3). \quad \approx 170.000$$

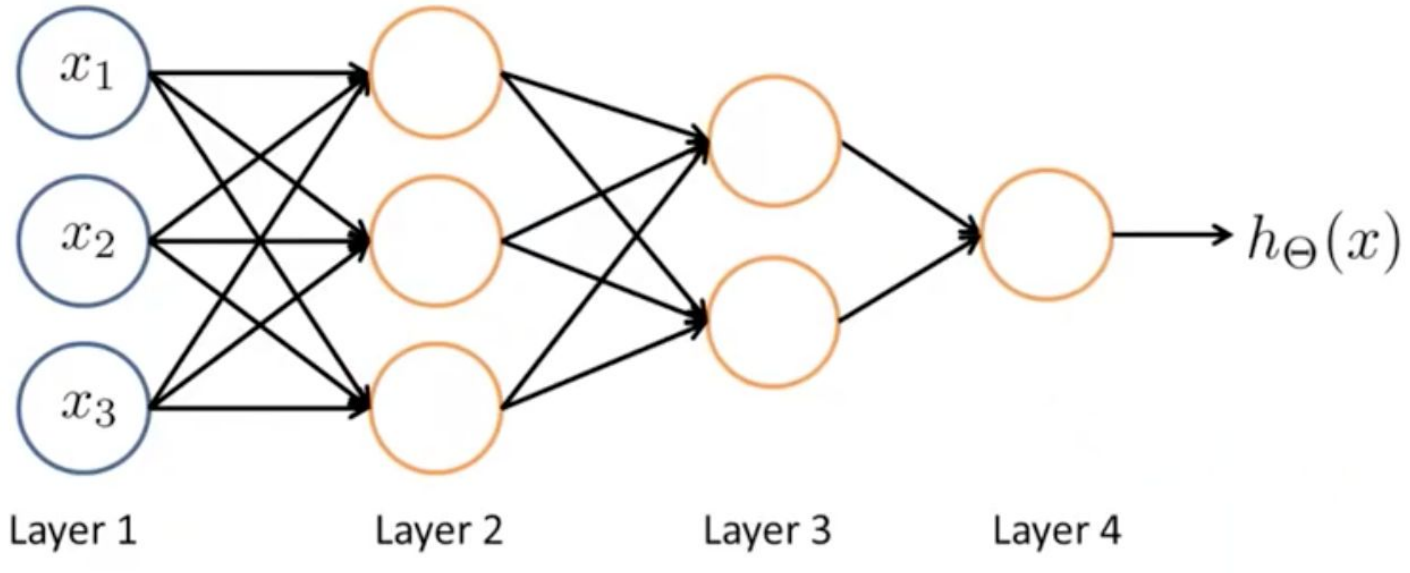
$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \theta_6 x_1 x_2^2 + \cdots)$$

$$x_1^2, x_1 x_2, x_1 x_3, \ldots, x_1 x_{100}, x_2^2, x_2 x_3, \ldots \approx 5000\ features.$$

$$\approx \frac{n^2}{2}$$

$$x_1 x_2 x_3, x_1 x_3 x_4, \ldots, x_1 x_{99} x_{100}, x_2 x_3 x_4, \ldots \approx O(n^3). \quad \approx 170.000$$

OTHER NETWORK ARCHITECTURES

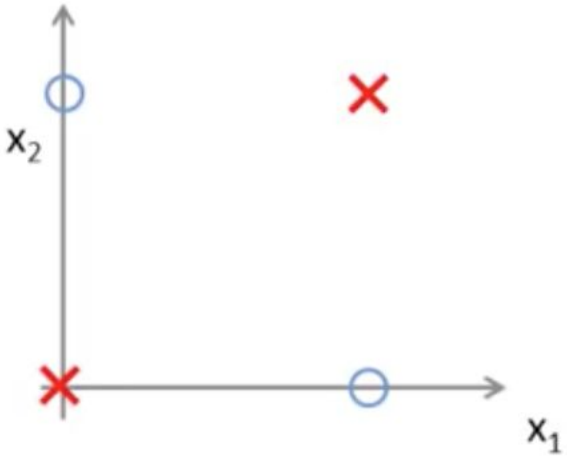


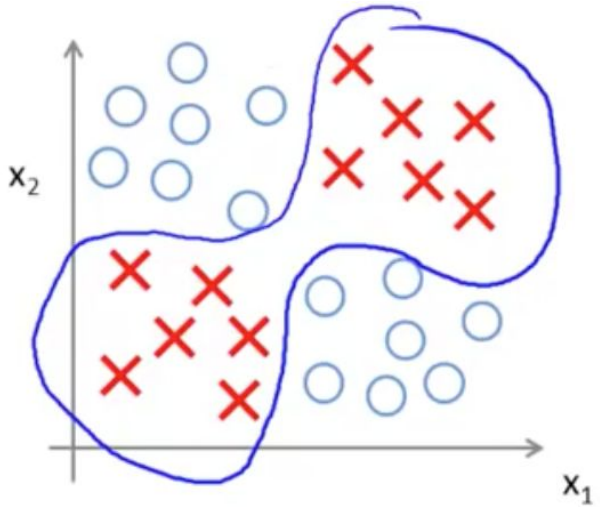
Layer 1 – Input layer

Layer 2, 3 – Hidden layer

Layer 4 – Output layer

NON-LINEAR CLASSIFICATION EXAMPLE: XOR/XNOR





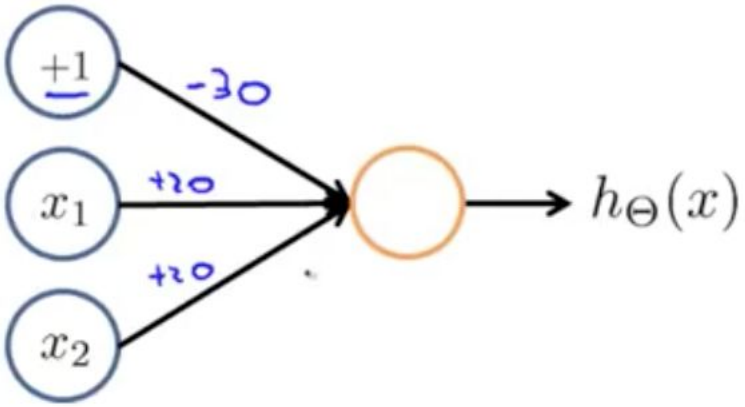
$$+ \theta_5 x_2^2 + \theta_6 x_1 x_2^2 + \cdots)$$

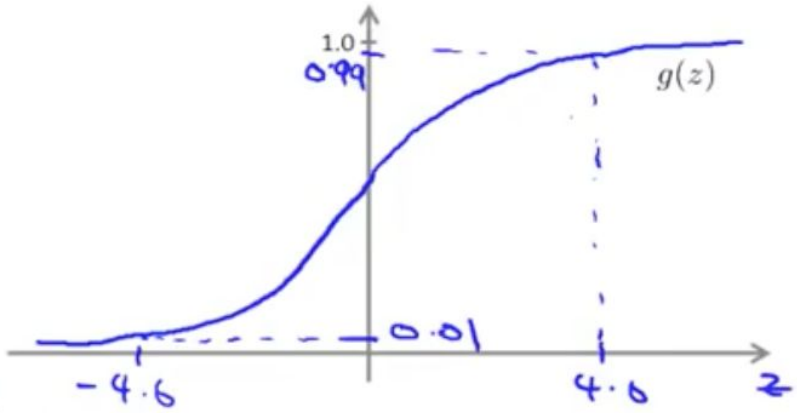
$$x_1^2, x_1x_2, x_1x_3, \ldots, x_1x_{100}, x_2^2, x_2x_3, \ldots$$
$$\approx 5000\ features.$$
$$\approx \frac{n^2}{2}$$

SIMPLE EXAMPLE: AND

$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \theta_6 x_1 x_2^2 + \cdots)$

$x_1^2, x_1 x_2, x_1 x_3, \ldots, x_1 x_{100}, x_2^2, x_2 x_3, \ldots$
$\approx 5000\ features.$
$\approx \frac{n^2}{2}$



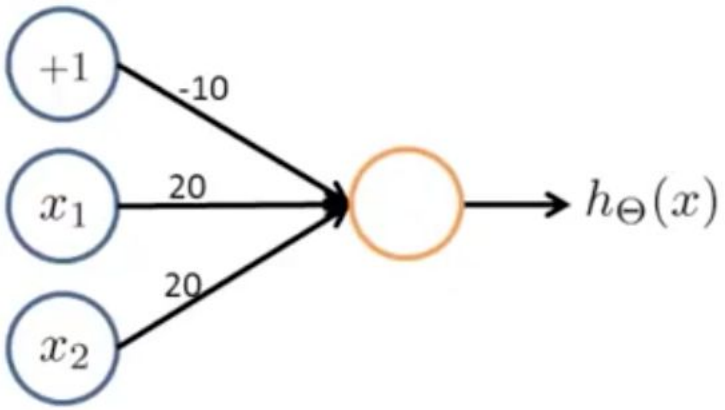


x_1	x_2	$h_\Theta(x)$
0	0	$g(-30) \approx 0$
0	1	$g(-10) \approx 0$
1	0	$g(-10) \approx 0$
1	1	$g(10) \approx 1$

$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \theta_6 x_1 x_2^2 + \cdots)$

$x_1^2, x_1 x_2, x_1 x_3, \ldots, x_1 x_{100}, x_2^2, x_2 x_3, \ldots$
$\approx 5000\ features.$
$\approx \frac{n^2}{2}$

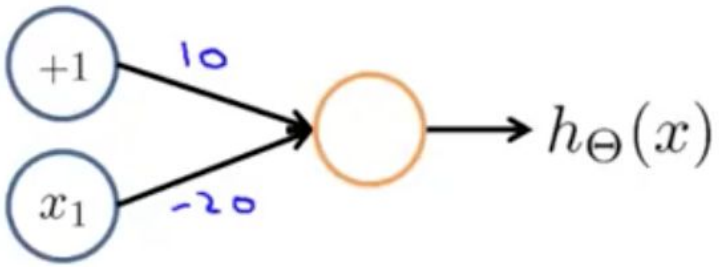
EXAMPLE: OR FUNCTION



$g(-10 + 20x_1 + 20x_2)$

x_1	x_2	$h_\Theta(x)$
0	0	$g(-10) \approx 0$
0	1	$g(10) \approx 1$
1	0	≈ 1
1	1	≈ 1

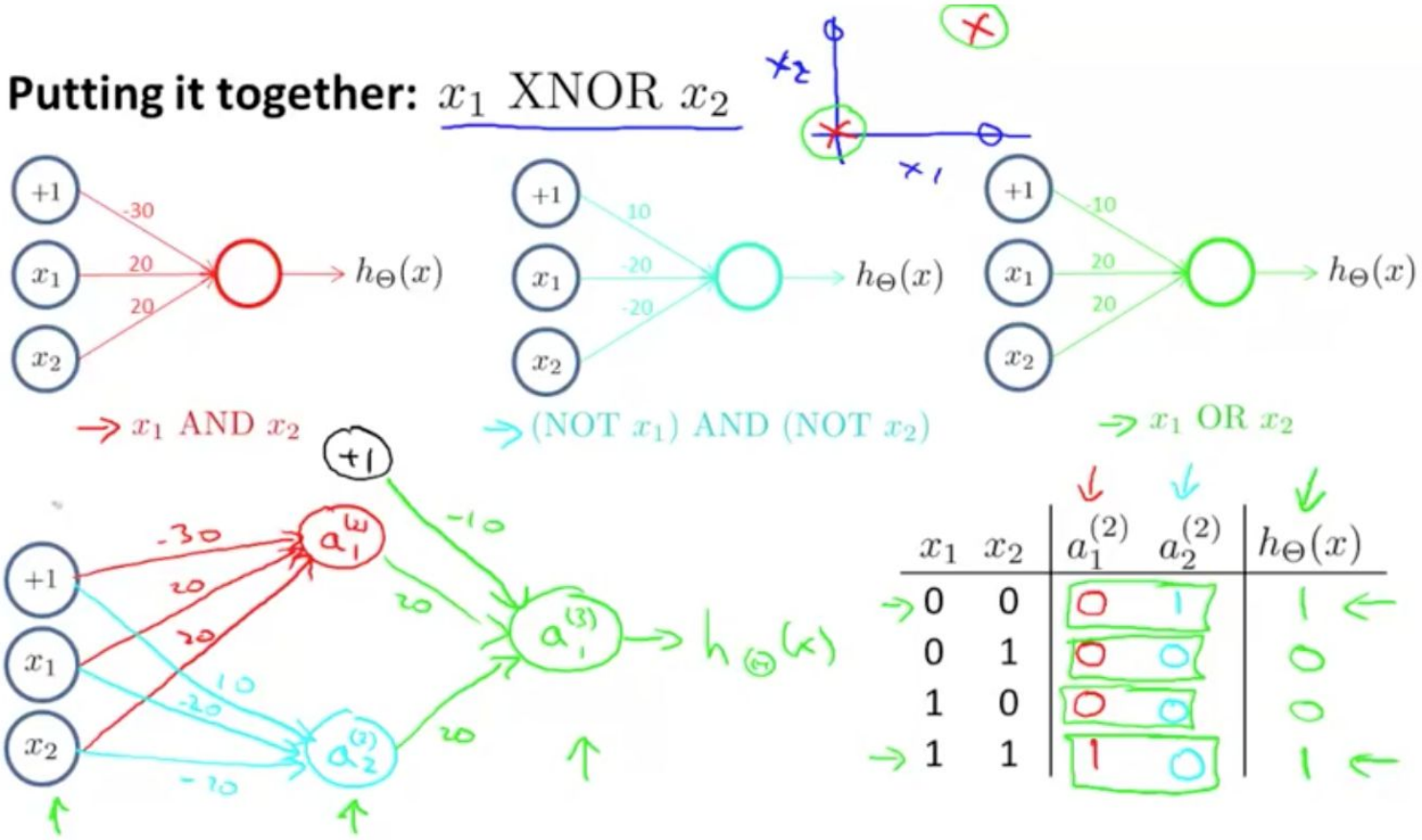
EXAMPLE: NOT FUNCTION



x_1	$h_\Theta(x)$
0	$g(10) \approx 1$
1	$g(-10) \approx 0$

$$h_\Theta(x) = g(10 - 20x_1)$$

EXAMPLE: XOR FUNCTION



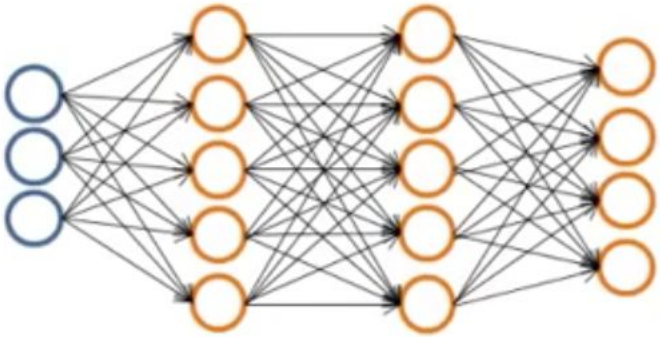
MULTICLASS CLASSIFICATION

Pedestrian

Car

Motorcycle

Truck

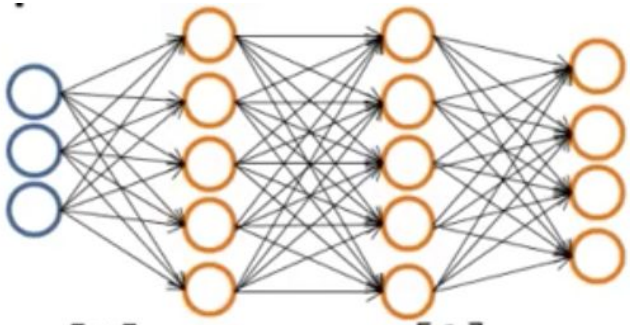


$h_\Theta(x) \in \mathbb{R}^4$

Want $h_\Theta(x) \approx \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$, $h_\Theta(x) \approx \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}$, $h_\Theta(x) \approx \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$, etc.

when pedestrian when car when motorcycle

MULTIPLE OUTPUT UNITS: ONE-VS-ALL



$h_\Theta(x) \in \mathbb{R}^4$

Want $h_\Theta(x) \approx \begin{bmatrix}1\\0\\0\\0\end{bmatrix}$, $h_\Theta(x) \approx \begin{bmatrix}0\\1\\0\\0\end{bmatrix}$, $h_\Theta(x) \approx \begin{bmatrix}0\\0\\1\\0\end{bmatrix}$, etc.

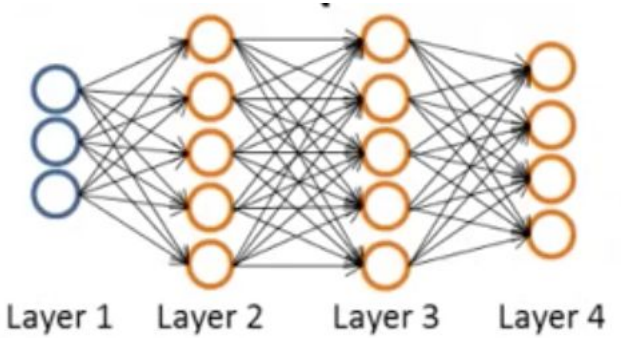
when pedestrian when car when motorcycle

Training set: $(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(m)}, y^{(m)})$

$y^{(i)}$ one of $\begin{bmatrix}1\\0\\0\\0\end{bmatrix}$, $\begin{bmatrix}0\\1\\0\\0\end{bmatrix}$, $\begin{bmatrix}0\\0\\1\\0\end{bmatrix}$, $\begin{bmatrix}0\\0\\0\\1\end{bmatrix}$

pedestrian car motorcycle truck

NEURAL NETWORK LEARNING COST FUNCTION



$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \theta_6 x_1 x_2^2 + \cdots)$$

$$x_1^2, x_1 x_2, x_1 x_3, \ldots, x_1 x_{100}, x_2^2, x_2 x_3, \ldots \approx 5000\ features. \approx \frac{n^2}{2}$$

$$x_1 x_2 x_3, x_1 x_3 x_4, \ldots, x_1 x_{99} x_{100}, x_2 x_3 x_4, \ldots \approx O(n^3). \quad \approx 170.000$$

$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \theta_6 x_1 x_2^2 + \cdots)$$

$$x_1^2, x_1 x_2, x_1 x_3, \ldots, x_1 x_{100}, x_2^2, x_2 x_3, \ldots \approx 5000\ features. \approx \frac{n^2}{2}$$

$$x_1 x_2 x_3, x_1 x_3 x_4, \ldots, x_1 x_{99} x_{100}, x_2 x_3 x_4, \ldots \approx O(n^3). \quad \approx 170.000$$

NEURAL NETWORK LEARNING
COST FUNCTION

$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \theta_6 x_1 x_2^2 + \cdots)$$

$$x_1^2, x_1 x_2, x_1 x_3, \ldots, x_1 x_{100}, x_2^2, x_2 x_3, \ldots$$
$$\approx 5000\ features.$$
$$\approx \frac{n^2}{2}$$

$$x_1 x_2 x_3, x_1 x_3 x_4, \ldots, x_1 x_{99} x_{100}, x_2 x_3 x_4, \ldots$$
$$\approx O(n^3). \qquad \approx 170.000$$

BACKPROPAGATION ALGORITHM: GRADIENT COMPUTATION

$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1 x_2 + \theta_4 x_1^2 + \theta_5 x_2^2 + \theta_6 x_1 x_2^2 + \cdots)$$

$$x_1^2, x_1 x_2, x_1 x_3, \ldots, x_1 x_{100}, x_2^2, x_2 x_3, \ldots \approx 5000\ features.$$

$$\approx \frac{n^2}{2}$$

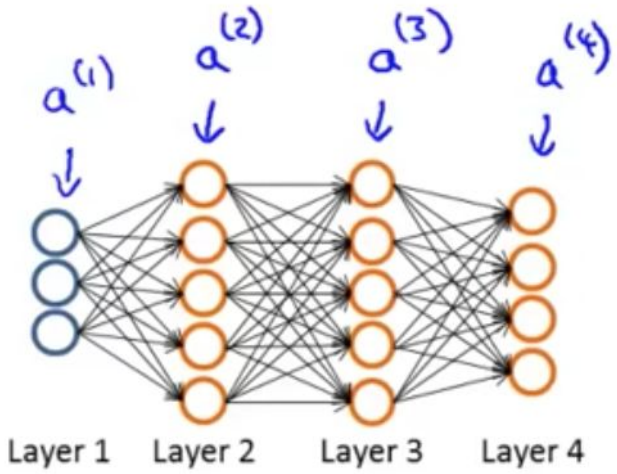
$$x_1 x_2 x_3, x_1 x_3 x_4, \ldots, x_1 x_{99} x_{100}, x_2 x_3 x_4, \ldots \approx O(n^3). \qquad \approx 170.000$$

BACKPROPAGATION ALGORITHM: GRADIENT COMPUTATION

$$g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_1 x_2 + \theta_5 x_2^2 + \theta_6 x_1 x_2^2 + \cdots)$$

$$x_1^2, x_1 x_2, x_1 x_3, \ldots, x_1 x_{10} \approx 5000 \; features.$$

$$\approx \frac{n^2}{2}$$



$$x_1 x_2 x_3, x_1 x_3 x_4, \ldots, x_1 x_{99} x_{100}, x_2 x_3 x_4, \ldots$$

$$\approx O(n^3). \qquad \approx 170.000$$

THANK YOU!